

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re the Application of:

KEON-HEO CHA, ET AL.

For: **INFORMATION GENERATION AND
RETRIEVAL METHOD BASED ON
STANDARDIZED FORMAT OF SENTENCE
STRUCTURE AND SEMANTIC STRUCTURE
AND SYSTEM USING THE SAME - UTILITY**

Honorable Commissioner of
Patents and Trademarks
Washington, D.C. 20231

Request for Priority

Sir:

Applicant respectfully requests a convention priority for the above-captioned application, namely
Korean application number 2000-74768 filed December 8, 2000.

☒ A certified copy of the document is being submitted herewith.

Respectfully submitted,

BLAKELY, SOKOLOFF, TAYLOR & ZAFMAN

Dated: _____

5/8/01


Eric S. Hyman, Reg. No. 30,139

12400 Wilshire Blvd., 7th Floor
Los Angeles, California 90025
Telephone: (310) 207-3800

J1046 U.S. PTO
09/852317
05/08/01
#3

<Priority Document Translation>

THE KOREAN INDUSTRIAL
PROPERTY OFFICE

This is to certify that the following application annexed
hereto is a true copy from the records of the Korean Industrial
Property Office.

Application Number : 2000-74768 (Patent)

Date of Application : December 8, 2000

Applicant(s) : ELECTRONICS AND TELECOMMUNICATIONS
RESEARCH INSTITUTE

March 6, 2001

COMMISSIONER



대한민국 특허청

KOREAN INTELLECTUAL PROPERTY OFFICE

별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Intellectual
Property Office.

출원 번호 : 특허출원 2000년 제 74768 호
Application Number

출원 년 월 일 : 2000년 12월 08일
Date of Application

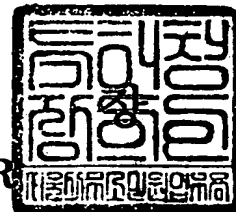
출원 인 : 한국전자통신연구원
Applicant(s)



2001 년 03 월 06 일

특 허 청

COMMISSIONER



【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0001
【제출일자】	2000. 12. 08
【발명의 명칭】	표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성 검색 장치 및 그 방법
【발명의 영문명칭】	Method of information generation and retrieval system based on a standardized Representation format of sentences structures and meanings
【출원인】	
【명칭】	한국전자통신연구원
【출원인코드】	3-1998-007763-8
【대리인】	
【성명】	특허법인 신성 정지원
【대리인코드】	9-2000-000292-3
【포괄위임등록번호】	2000-051975-8
【대리인】	
【성명】	특허법인 신성 원석희
【대리인코드】	9-1998-000444-1
【포괄위임등록번호】	2000-051975-8
【대리인】	
【성명】	특허법인 신성 박해천
【대리인코드】	9-1998-000223-4
【포괄위임등록번호】	2000-051975-8
【발명자】	
【성명의 국문표기】	차건희
【성명의 영문표기】	CHA, Keon Hoe
【주민등록번호】	690213-1580911
【우편번호】	302-223
【주소】	대전광역시 서구 탄방동 주공아파트 103-107
【국적】	KR

【발명자】

【성명의 국문표기】 정의석
 【성명의 영문표기】 CHUNG, Eui Sok
 【주민등록번호】 730203-1347617
 【우편번호】 305-345
 【주소】 대전광역시 유성구 신성동 208-3 은지빌라 101
 【국적】 KR

【발명자】

【성명의 국문표기】 임수중
 【성명의 영문표기】 LIM, Soo Jung
 【주민등록번호】 710919-1536412
 【우편번호】 133-093
 【주소】 서울특별시 성동구 금호동3가 1143번지 2층
 【국적】 KR

【발명자】

【성명의 국문표기】 강현규
 【성명의 영문표기】 KANG, Hyun Kyu
 【주민등록번호】 620805-1457258
 【우편번호】 305-350
 【주소】 대전광역시 유성구 가정동 236-2 KIT교수아파트 14-404
 【국적】 KR

【심사청구】

청구

【취지】

특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인
 특허법인 신성 정지원 (인) 대리인
 특허법인 신성 원석희 (인) 대리인
 특허법인 신성 박해천 (인)

【수수료】

【기본출원료】	20 면	29,000 원
【가산출원료】	30 면	30,000 원
【우선권주장료】	0 건	0 원
【심사청구료】	22 항	813,000 원

【합계】	872,000 원
【감면사유】	정부출연연구기관
【감면후 수수료】	436,000 원
【첨부서류】	1. 요약서·명세서(도면)_1통

【요약서】

【요약】

본 발명은 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것으로, 사용자 자신이 제공하고자 하는 정보를 기술하거나 또는 자신이 원하는 정보를 검색할 수 있도록 하기 위하여, 표준화된 자연언어 구문구조 및 의미구조 표현 기법을 이용하여 정보를 생성 및 저장하여 검색할 수 있도록 하기 위한 정보 생성/검색 장치 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공하기 위하여, 정보 공급자에 의해 기술되는 자연언어 문장(정보 및 지식)을 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환하여 색인하는 제 1 단계; 및 사용자로부터 입력되는 자연언어 질의어 문장을 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환하여 상기 색인된 정보중에서 사용자의 요구사항에 적합한 정보를 검색하여 제공하는 제 2 단계를 포함하며, 정보거래와 관련된 정보 생성 및 검색 장치 등에 이용됨.

【대표도】

도 1

【색인어】

자연언어 문장, 구문구조 표준, 의미구조 표준, 분할 그래프

【명세서】**【발명의 명칭】**

표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치 및 그 방법
{Method of information generation and retrieval system based on a standardized
Representation format of sentences structures and meanings}

【도면의 간단한 설명】

도 1 은 본 발명에 따른 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/
검색 장치의 일실시에 구성도.

도 2a 및 도 2b 는 본 발명에 따른 표준화된 문장 구문구조 및 의미구조에 기반한
정보 생성/검색 방법에 대한 일실시에 흐름도.

도 3 은 도 2a 및 도 2b의 정보 생성/검색 방법 중 의미구조 생성 과정에 대한 일
실시에 상세흐름도.

도 4a 및 도 4b 는 본 발명에 따른 구문 범주와 그에 따른 변환 규칙 및 개념 처리
결과 생성된 서브 그래프에 대한 예시도.

도 5a 내지 도 5f는 본 발명에 따른 분할 그래프의 색인 결과 생성된 테이블 구조
와 레코드 내용을 설명한 예시도.

도 6 은 본 발명에 따른 사용자의 요구 정보 및 문서 추출 과정에 대한 일실시에
흐름도.

* 도면의 주요 부분에 대한 부호의 설명

11 : 데이터 저장부 12 : 정보 제공 인터페이스

13 : 입력 문장 분석부 14 : 대화 처리부

15 : 의미구조 처리부 16 : 정보 수요 인터페이스

【발명의 상세한 설명】

【발명의 목적】

【발명이 속하는 기술분야 및 그 분야의 종래기술】

<11> 본 발명은 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것으로, 특히 인터넷상에서 빠르고 효율적으로 정보 및 지식을 생성, 검색 및 유통하도록 하기 위하여, 구문구조 및 의미구조의 표준(standard)에 적합한 자연언어 문장을 생성하여 정보를 기술하고 검색할 수 있도록 하는, 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체에 관한 것이다.

<12> 정보화 사회로의 이행에 가장 관건이 되는 것은 대규모의 자료를 저장, 관리하고 이를 원활히 유통할 수 있는 정보서비스의 확립에 있다. 모든 정보자료들이 디지털화되어 데이터베이스에 저장되고 네트워크에 연결되어 온라인으로 어느 지역에서나 원격으로 접근하고 이용할 수 있게 됨에 따라 정보관리와 정보검색은 중요

한 핵심 기술의 하나로 여겨지게 되었다. 정보검색은 정보를 축적하고 가공하는 정보가 공 과정과 정보 요구자에게 적절한 정보를 제시하고 탐색해 주는 정보제시의 부분으로 크게 나뉜다. 또한 정보 요구자들로 하여금 정보서비스를 보다 편리하고 효율적으로 사용하도록 도와주는 사용자 인터페이스 기술도 이에 포함된다.

<13> 정보거래(information brokerage)와 관련해서는 미국의 경우 아메리칸 온 라인 ((AOL : American On-Line)과 같은 다국적 PC 통신업체에서 각종 정보 유통서비스를 제공하고 있으며, 웹정보 검색엔진 업체(예를 들면, Yahoo!)와 제휴하여 그 사업 영역을 넓히고 있는 중이지만 문장의 의미표현 및 의미유사도에 기반하여 적합한 정보거래 후보를 정확하고 신속하게 알선하며 거래를 중개해주는 기술은 아직 개발되지 않고 있다.

<14> 의미 표현 기술과 관련해서는 1998년부터 미국 메사추세츠 대학교의 지능형 정보 검색 센터(CIIR: Center for Intelligent Information Retrieval)과 같은 정보검색 및 자연언어처리 연구소에서 어휘개념 및 정보의 의미분석 기반의 정교한 정보 추출 및 정보 검색 기술을 활용한 응용 기술을 개발하려는 움직임을 보이고 있으나 정보거래에 활용하려는 움직임은 없다. 또한, 미국 메릴랜드대와 남가주대에서 온탈로지 명세 언어 (OML: Ontology Markup Language), 개념지식 명세 언어(CKML: Conceptual Knowledge Markup Language) 등을 이용한 웹페이지 마크업 언어가 제안되고 있으며, 뉴멕시코 주립 대학교의 'Mikro-Kosmos', 프린스턴대의 'WordNet', 싸이코프(Cycorp)의 'CYC' 및 KIF(Knowledge Interchange Format)와 온탈로지 공유 재사용에 관한 연구가 진행되고 있고, 각 기관은 그들의 기법을 표준

안으로 제안하려는 움직임을 보이고 있으며, 최근에는 국제전기전자기술자협회(IEEE : Institute of Electrical and Electronics Engineers)에서 SUO(Standard Upper Ontology)라는 표준화된 의미 계층구조를 통해 일반적이고 추상적인 상위 수준에서의 개념을 정립하려는 노력을 기울이고 있다.

<15> 정보거래와 관련된 국내 기술 현황으로는 천리안, 나우누리, 하이텔 등 PC 통신 업체에서는 유망한 영역의 정보제공 사업 제안자의 사업능력을 판정하여 정보 제공자(IP) 사업권을 주어, 독점적으로 정보를 수집 가공하여 PC 통신 방에 적절한 메뉴와 정보를 등록하도록 하고, 사용자에게서 이 정보를 열람하는 시간동안의 사용료를 받는 방식으로 이루어지고 있다. 또한, 정보 거래 센터라는 업체가 1997년 말에 발족하여 인터넷을 통하여 정보 거래를 취급하려고 하고 있으며, 특히 기술 분야의 중개를 시작하였다. 그러나, 정보 거래의 프로토콜이 마련되지 않았고, 적합한 거래 후보를 자동으로 정확하게 선택하는 개념유사도 기반 정보 추출/매칭 기술이 적용되지 않아 대규모의 다양한 정보의 거래를 처리하기는 역부족인 상황이다.

<16> 의미 표현 기술과 관련해서 국내에서는 의미표현 및 관련 연구는 전북대, 포항공대 등을 중심으로 개념 그래프(Conceptual Graph)에 대한 기초 연구가 진행중이며, 개념분류체계에 대한 연구는 에트리(ETRI), 울산대, 한남대, 인천대, 호서대 등에서 진행되고 있으나 국제적인 추세에 맞추어 의미표현 및 관련 기술의 표준화에 대한 노력은 이루어지지 못하고 있는 실정이다.

<17> 따라서, 구문구조 및 의미구조의 표준을 연구하는 기술분야에서는 인터넷상

에서 빠르고 효율적인 정보 및 지식을 생성/검색할 수 있도록 구문구조 및 의미구조의 표준(standard)에 적합한 자연언어 문장을 생성하여 정보를 기술하고 검색할 수 있는 방안이 필수적으로 요구되고 있다.

【발명이 이루고자 하는 기술적 과제】

<18> 본 발명은, 상기한 바와 같은 요구에 부응하기 위하여 안출된 것으로, 사용자 자신이 제공하고자 하는 정보를 기술(description)하거나 또는 자신이 원하는 정보를 검색할 수 있도록 하기 위하여, 표준화된 자연언어 구문구조 및 의미구조 표현 기법 (standardized formats for the structures and meanings of natural language sentences)을 이용하여 정보를 생성 및 저장하여 검색할 수 있도록 하기 위한 정보 생성/검색 장치 및 그 방법과 상기 방법을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공하는데 그 목적이 있다.

【발명의 구성 및 작용】

<19> 상기 목적을 달성하기 위한 본 발명은, 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치에 있어서, 사용자로부터의 정보 제공을 위한 문장이나 정보 요구를 위한 질의문을 분석하기 위해 사용하는 언어 지식 데이터와 문장의 의미를 개념 그래프로 나타내는 의미 표현 데이터와 웹 문서를 저장하기 위한 데이터 저장수단; 상기 사용자로부터 정보 제공을 위한 자연언어 문장의 생성이나 정보의 요구 명세를 위한 자연어 질의 문장을 입력받기 위한 입력수단; 상기 데이터 저장수단의 데이터를 참조하여

상기 사용자로부터 입력된 자연언어 문장 또는 자연언어 질의 문장의 구문구조를 분석하여 의미구조를 생성하기 위한 입력 문장 분석수단; 상기 입력 문장 분석수단을 통해 분석된 의미 구조를 분할하여 데이터베이스에 색인하여 저장하거나 의미 적합도를 계산하여 요구 정보 명세에 가장 의미적으로 적합한 제공 정보 및 문서를 검색하기 위한 의미 구조 처리수단; 상기 입력 문장 분석수단으로부터 제시된 오류 데이터를 구문구조 표준과 의미구조 표준에 따라 수정한 문장 형식 규칙, 상기 색인/검색 결과를 출력하기 위한 대화형 처리수단; 및 상기 대화형 처리수단으로부터 출력되는 데이터를 사용자에게 제공하기 위한 정보 제공수단을 포함하여 이루어진 것을 특징으로 한다.

<20> 또한, 본 발명은, 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치에 적용되는 정보 생성/검색 방법에 있어서, 정보 공급자에 의해 기술되는 자연언어 문장(정보 및 지식)을 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환하여 색인하는 제 1 단계; 및 사용자로부터 입력되는 자연언어 질의어 문장을 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환하여 상기 색인된 정보중에서 사용자의 요구사항에 적합한 정보를 검색하여 제공하는 제 2 단계를 포함하여 이루어진 것을 특징으로 한다.

<21> 또한, 본 발명은, 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성 장치에 적용되는 정보 생성 방법에 있어서, 상기 정보 공급자에 의해 입력되는 자연언어 문장을 구문구조 및 의미구조의 표준에 따라 구문구조 및 의미구조의 중의성이 해소된 문장으로 생성하는 제 1 단계; 상기 생성된 문장을 구문분석과 의미분석을 통하여 개념 그래프로 변환하는 제 2 단계; 및 상기 변환된 개념 그래프를 관계 노드에 의한 테이블의 레코드 값으로 변환하여 색인하는 제 3 단계를 포함하여 이루어진 것을 특징으로 한다.

<22> 또한, 본 발명은, 표준화된 문장 구문구조 및 의미구조에 기반한 정보 검색 장치에 적용되는 정보 검색 방법에 있어서, 상기 사용자로부터 자연언어 질의어 문장을 입력받아 구문구조 및 의미구조를 분석하여 개념 그래프로 변환하는 제 1 단계; 상기 질의어를 개념 그래프와 의미적으로 가장 근접한 데이터베이스내의 개념 그래프를 탐색하여 의미 적합도를 계산하는 제 2 단계; 및 상기 탐색된 개념 그래프에 의해 색인된 정보를 추출하여 상기 사용자에게 제공하는 제 3 단계를 포함하여 이루어진 것을 특징으로 한다.

<23> 또한, 본 발명은, 프로세서를 구비한 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치에, 정보 공급자에 의해 기술되는 자연언어 문장(정보 및 지식)을 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환하여 색인하는 제 1 기능; 및 사용자로부터 입력되는 자연언어 질의어 문장을 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환하여 상기 색인된 정보중에서 사용자의 요구사항에 적합한 정보를 검색하여 제공하는 제 2 기능을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

<24> 또한, 본 발명은, 프로세서를 구비한 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성 장치에, 상기 정보 공급자에 의해 입력되는 자연언어 문장을 구문구조 및 의미구조의 표준에 따라 구문구조 및 의미구조의 중의성이 해소된 문장으로 생성하는 제 1 기능; 상기 생성된 문장을 구문분석과 의미분석을 통하여 개념 그래프로 변환하는 제 2 기능; 및 상기 변환된 개념 그래프를 관계 노드에 의한 테이블의 레코드 값으로 변환하여 색인하는 제 3 기능을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

<25> 또한, 본 발명은, 프로세서를 구비한 표준화된 문장 구문구조 및 의미구조에 기반

한 정보 검색 장치에, 상기 사용자로부터 자연언어 질의어 문장을 입력받아 구문구조 및 의미구조를 분석하여 개념 그래프로 변환하는 제 1 기능; 상기 질의어를 개념 그래프와 의미적으로 가장 근접한 데이터베이스내의 개념 그래프를 탐색하여 의미 적합도를 계산하는 제 2 기능; 및 상기 탐색된 개념 그래프에 의해 색인된 정보를 추출하여 상기 사용자에게 제공하는 제 3 기능을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

<26> 본 발명은, 구문구조 및 의미구조 표준에 따라 자연언어 문장을 작성함으로써, 사용자에게 의해 기술되는 정보 및 지식을 개념 그래프로 변환하여 저장하고 색인하는 '정보 제공'(information supply) 과정과 사용자의 질의어를 분석하여 사용자의 요구사항에 적합한 정보를 검색하고 추출하는 '정보 요구' 과정을 포함한다.

<27> 상기 '정보 제공' 과정은 구문구조 및 의미구조의 표준에 따라 정보 및 지식이 생성되도록 돕는 정보 기술 지원 도구(a toolset for supporting information specification)에 의해 구문구조 및 의미구조의 중의성(ambiguities)이 해소된 문장을 생성하고, 생성된 문장을 구문 분석과 의미분석을 통하여 개념 그래프(conceptual graph)로 변환하며, 개념그래프를 관계 노드(relation node)에 의한 테이블의 레코드 값으로 변환하여 저장하고 색인한다.

<28> 상기 '정보 요구' (information request) 과정은 사용자의 자연언어 질의어를 구문/의미분석을 통하여 개념 그래프로 변환하며, 질의어의 개념 그래프와 의미적으로 가장 근접한 데이터베이스내의 개념 그래프를 탐색하여 의미 적합도(semantic appropriateness degree) 계산을 수행하며, 탐색된 개념 그래프에 의해 색인된 정보를 사용자에게 보여주는 것을 특징으로 한다.

- <29> 즉, 본 발명은, 구문구조 및 의미구조의 표준에 의한 효율적인 정보 거래를 위하여 사용자의 정보나 지식이 생성되는 단계에서부터 사용자와의 상호 작용을 통해 구조적/의미적 중의성을 해결하고 사용자의 명세 정보가 상기 구문구조 및 의미구조 표준에 일치하도록 함으로써, 정보와 지식의 검색과 유통이 효율적이고 정확하게 이루어지도록 하는 것을 특징으로 한다.
- <30> 상술한 목적, 특징들 및 장점은 첨부된 도면과 관련한 다음의 상세한 설명을 통하여 보다 분명해 질 것이다. 이하, 첨부된 도면을 참조하여 본 발명에 따른 바람직한 일 실시예를 상세히 설명한다.
- <31> 도 1 은 본 발명에 따른 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치의 일 실시예 구성도이다.
- <32> 도 1에 도시된 바와 같이, 본 발명에 따른 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치는, 사용자로부터의 정보 제공을 위한 문장이나 정보 요구를 위한 질의문을 분석하기 위해 사용하는 언어 지식 데이터와 문장의 의미를 개념 그래프로 나타내는 의미 표현 데이터와 웹 문서를 저장하기 위한 데이터 저장부(11)와, 상기 사용자로부터 거래를 목적으로 하는 정보 제공을 위한 문장의 생성이나 정보의 요구 명세를 위한 자연어 질의 문장을 입력받기 위한 정보 제공 인터페이스(12)와, 데이터 저장부(11)의 데이터를 참조하여 상기 사용자로부터 입력된 자연어 문장을 분석하여 의미구조를 생성하기 위한 입력 문장 분석부(13)와, 입력 문장 분석부(13)를 통해 사용자가 작성한 문장에서 발생하는 철자나 띄어쓰기 오류, 구문구조나 의미구조의 표준과 위배되는 비문, 구문구조의 중의성 및 의미구조의 중의성 등이 발생했을 때 이에 대한 의미를 해결하여 그 결과와 색인 및 검색 결과를 출력하기 위한 대화 처리부(14)와, 입력 문장 분

석부(13)로부터 의미구조를 입력받아 색인/저장하거나 검색하기 위한 의미구조 처리부(15) 및 대화 처리부(14)부터 출력되는 데이터를 사용자에게 제공하기 위한 정보 수요 인터페이스(16)를 구비한다.

<33> 여기서, 데이터 저장부(data storage part)(11)는 사전(lexicon) 저장장치(111), 용언 격 프레임(predicate case frame) 저장장치(112), 명사 개념 분류망(noun thesaurus) 저장장치(113), 개념 그래프 데이터베이스(114) 및 웹 문서 데이터베이스(115)로 이루지고, 입력 문장 분석부(13)는 형태소 분석기(morphological analyzer)(131), 구문 분석기(parser)(132) 및 의미구조 생성기(semantic structure generator)(133)로 구성되고, 대화 처리부(interactive response processor)(14)는 분석 실패 처리기(analysis failure processor)(141), 응답 생성기(response generator)(142)로 구성되며, 의미구조 처리부(semantic structure processor)(15)는 개념 그래프 변환기(conceptual graph transformer)(151), 개념 그래프 색인기(conceptual graph indexer)(152), 개념 그래프 검색기(conceptual graph searcher)(107)(153)로 이루어진다.

<34> 상기와 갖는 구조를 갖는 본 발명의 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치의 동작을 상세하게 설명하면 다음과 같다.

<35> 먼저, 사용자로부터 자연언어 문장을 통해 제공하고자 하는 정보를 입력받는 정보 제공 인터페이스(12)나 제공받고자 하는 정보에 대한 명세를 입력받는 정보 수요 인터페이스(16)가 입력 문장을 입력 문장 분석부(13)로 전달하면, 입력문장 분석부(13)는 사용자로부터 입력된 문장을 받아 데이터 저장부(11)에 있는 사전 저장 장치(111), 용언 격 프레임 저장 장치(112) 및 명사 개념 분류망 저장 장치(113)의 데이터와 비교하여 순차

적으로 분석을 하는데, 이때 형태소 분석기(141)에서 형태소 분석을 수행하고, 구문 분석기(132)에서 구문 분석을 수행하여 구문구조 트리를 생성하고, 의미구조 생성기(133)에서 의미 분석을 수행하여 의미구조를 생성한다.

<36> 의미구조 처리부(15)의 개념 그래프 변환기(151)는 의미구조 생성기(133)로부터 출력된 개념 그래프를 의미 관계에 따라 분리해내고 개념 그래프 색인기(153)는 이렇게 분리된 각각의 관계로 연결된 개념의 쌍의 레코드에 의해 사용자의 제공 정보가 있는 원래의 웹문서에 대한 색인을 수행하고 개념 그래프 검색기(152)는 사용자 질의어의 의미구조와 저장된 의미구조들 사이에서 의미 적합도(semantic relevance)를 계산하여 적합도가 높은 제공 정보를 탐색한다.

<37> 대화 처리부(14)의 분석 실패 처리기(141)는 형태소 분석기(131), 구문 분석기(132), 의미구조 생성기(133)가 분석 과정에서 사용자의 표준 규칙 위반에 의해 발생하는 분석 실패에 대해 생성하는 오류 유형 데이터를 입력받아 구문구조 표준 및 의미구조 표준에 일치하는 수정된 문장 형식 규칙을 찾아내고, 응답 생성기(142)는 분석 실패 처리기(141)로부터 입력되는 수정 문장 규칙이나 개념 그래프 검색기(152)로부터 입력되는 검색 결과나 개념 그래프 색인기(153)로부터 입력되는 색인 결과를 입력 받아 사용자에게 대한 응답 형식과 내용을 생성한다.

<38> 상기의 구문구조 표준과 의미구조 표준을 나타내면 다음과 같다.

<39> (1) 구문구조 표준

<40> 201. 논항

<41> NP1:(_NP+jc1)

- <42> NP2:(_NP+jc2)
- <43> NP3:(_NP+jc3)
- <44> 202. 기본문
- <45> S:(NP1 _VP)
- <46> S:(NP1 NP2 _VP)
- <47> S:(NP1 NP3 _VP)
- <48> S:(NP1 NP2 NP3 _VP)
- <49> 203. 명사구
- <50> NP:(nc|nn+[nc|nn]+[nc|nn])
- <51> NP:(nb|np)
- <52> 204. 관형사 확대 논항
- <53> _NP:(mm!..[dt]+[mm]!..[nu|ad] NP!-[nn,nb])
- <54> 205. 접속조사, 접속 부사의 논항 확대
- <55> _NP:(NP+jj NP)
- <56> _NP:(_NP maj _NP)
- <57> _NP:(NP!@+', ' maj NP)
- <58> 206. 속격의 논항 확대
- <59> _NP:(_NP+jm _NP!^[NP])
- <60> 207. 서술어
- <61> VP:(pv+[ep]+[ef])

- <62> VP:(pa+[ep]+[ef])
- <63> VP:(NP+co+[ep]+[ef])
- <64> VP:(NP+xsv+[ep]+[ef])
- <65> 208. 부사의 서술어 확대
- <66> VP:(mag _VP!^[VP])
- <67> VP:(maj _S)
- <68> 209. 보조용언의 서술어 확대
- <69> VP:(_VP:(_VP+ec px+[ec] [px])+[ep]+[ef])
- <70> 210. 관형절의 논항 확대
- <71> _NP:(ETMS:(_S+etm) NP)
- <72> 211. 명사절
- <73> _NP:(_S+etn)
- <74> _NP:(_S+etm '것')
- <75> 212. 인용절의 서술어 확대
- <76> VP:(JQTS:(_S+'라고|고|하고') _VP)
- <77> VP:(JQTS:(''+_S+''+ '라고|고|하고') _VP)
- <78> 213. 부사절의 서술어 확대
- <79> VP:(ADVS:(NP1 [NP2] [NP3] '게|듯이|도록') _VP)
- <80> VP:(ADVS:(NP1 [NP2] [NP3] '같이|없이|달리') _VP)
- <81> 214. 서술절

<82> VP: (NP1 _VP!^[VP])

<83> 215. 문의 접속

<84> S: (_S+ec _S)

<85> (2) 의미구조 표준

<86> <개념간 관계>

<87> 216. 행위자 관계(agent)

<88> '이/가', '은/는', '-께서', '-에서', '-의'

<89> 217. 경험자 관계(experiencer)

<90> '이/가', '은/는', '-의'

<91> 218. 대상 관계(object)

<92> '을/를', '-와/과', '-의'

<93> 219. 장소 관계(location)

<94> '-을', '-에', '-에서', '-[으]로'

<95> 219a. 시작점 관계(starting point)

<96> 219b. 지향점 관계(target point)

<97> 220. 시간 관계 : C1(TIME)C2

<98> '-에'

<99> 220a. 시작점 관계(starting point)

<100> 220b. 완료점 관계(target point)

- <101> 221. 목적 관계(purpose)
- <102> '-을', '-려면'
- <103> 222. 이유/원인 관계(causal-effect)
- <104> '-(으)로'
- <105> 223. 수용자 관계(receiver)
- <106> '-에 게', '-에'
- <107> 224. 결과 관계(result)
- <108> '-(으)로'
- <109> 225. 수단/방법 관계(instrument)
- <110> '-로', '-에', '-아/어[서]'
- <111> 226. 비교관계 : '-와/과', '-만큼', '-보다', '-처럼'
- <112> 226a. 일반비교 관계(comparison)
- <113> 226b. 동등비교 관계(equivalence)
- <114> 226c. 차등비교 관계(difference)
- <115> 226d. 비유 관계(analogy)
- <116> 227. 속성 관계(property)
- <117> '-의', '# ' [#은 공백]
- <118> 227a. 소유자 관계(owner)
- <119> 227b. 생산지/생산자 관계(source)
- <120> 227c. 소재지 관계(location)

- <121> 228d. 소속 관계(belonging to)
- <122> 229e. 제재 관계(substance)
- <123> 229f. 상호 관계(reciprocal)
- <124> 229g. 하위개념 관계(subsumption)
- <125> 229h. 부분-전체 관계(whole-part)
- <126> 229i. 속성/개체 관계(property of)
- <127> 229j. 동시/병치관계(juxtapose)
- <128> <개념 그래프간 관계>
- <129> 230. 병치 관계(Juxtaposition)
- <130> 231. 조건 관계(condition)
- <131> 232. 원인결과 관계(causal-effect)
- <132> 233. 전환 관계(context switch)
- <133> 233a. 선행 상황 미완(unfinished precedence)
- <134> 233b. 선행 상황 완료(finished precedence)
- <135> 234. 내포 관계(involvement)
- <136> 235. 가능 관계(possibility)
- <137> 236. 양보 관계(concession)
- <138> 다음으로, 상기 각 구성요소의 상세동작을 도 2a 내지 도 6을 참조하여 상세히 살펴보기로 한다.
- <139> 도 2a 및 도 2b 는 본 발명에 따른 표준화된 문장 구문구조 및 의미구조에 기반한

정보 생성/검색 방법에 대한 일실시에 흐름도이다.

- <140> 도 2a 및 도 2b에 도시된 바와 같이, 본 발명에 따른 표준화된 문장 구문구조 및 의미구조에 기반한 정보를 생성 및 검색하기 위하여, 정보 제공 인터페이스(16)를 통해 사용자 자신이 제공하고자 하는 정보나 제공받기를 원하는 정보에 대해서 자연어 문장을 입력하면(201), 입력된 자연어 문장이 형태소 분석기(131)에 의해 형태소 분석된다(202).
- <141> 이어서, 형태소 분석이 성공적으로 수행되었는지를 확인하여(203) 형태소 분석이 실패되었으면 형태소 분석 오류가 발생한 유형에 따라 오류 유형 데이터를 생성하고(204), 형태소 분석이 성공했으면 형태소 분석 결과를 가지고 구문 분석을 수행한다(205).
- <142> 이후, 정보 거래 표준에 적합한 구문구조가 생성되는지를 확인하여(206) 구문구조의 생성에 실패하면 구문 분석 오류 유형 데이터(parsing failure type data)를 생성하고(207), 구문구조의 생성에 성공하면 구문 분석 트리(parse tree)를 의미구조로 변환하여 의미구조를 생성한다(208).
- <143> 이어서, 의미구조 표준에 적합한 의미구조가 생성되는지를 확인하여(209) 의미구조 생성에 실패하면 의미구조 생성 오류 유형 데이터를 생성하고(210), 의미구조 생성에 성공하면 개념 그래프 변환기(151)에 입력하여 개념 그래프에 대한 분할(partition)을 이룬다(211).
- <144> 그리고, 현재 사용자의 입력 문장이 정보 제공용인지를 확인하여(212) 정보 제공용이면 그 문장의 분할된 개념 그래프를 각각의 의미 관계 노드(relation node)에 대해 색

인을 이루고(213), 개념 그래프 색인 결과에 따라 데이터를 생성한 후(214), 정보 거래 시스템의 응답의 형식과 내용을 생성하여(215) 최종적인 응답을 출력한다(216). 한편, 현재 사용자의 입력 문장이 정보 요구를 위한 질의어(query)라면 개념 그래프 데이터베이스(114)에 저장되어 있는 개념 그래프와 의미적으로 가장 가까운 개념 그래프를 검색하여(217) 개념 그래프 검색 결과를 생성한 후(218), 정보 거래 시스템의 응답의 형식과 내용을 생성하는 과정(215)을 진행한다.

<145> 한편, 형태소 분석 오류 유형 데이터 생성(204), 구문 분석 오류 유형 데이터 생성(207) 및 의미구조 생성 오류 유형 데이터를 생성한 후(210), 분석 실패 처리기(141)에 의해 오류 수정 규칙 데이터를 생성한 후(219), 정보 거래 시스템의 응답의 형식과 내용을 생성하는 과정(215)을 진행한다.

<146> 상기와 갖는 구조를 갖는 본 발명의 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법의 동작을 상세하게 설명하면 다음과 같다.

<147> 정보 거래를 위한 정보의 생성 및 검색 방법에 있어서, 먼저 텍스트 편집기를 통해 문장을 입력하고, 입력된 문장에 대해 형태소 분석을 수행하여 형태소 중의성이나 철자 오류 등이 발생할 경우 사용자에게 분석 결과를 제시하여 사용자로부터 수정된 문장을 입력받는다.

<148> 문장의 형태소 분석 결과를 입력받아 구문분석을 수행하여 구문구조 표준에 불일치하거나 구문구조의 중의성이 발생할 경우 사용자에게 분석 결과를 제시하고 사용자로 하여금 수정된 문장을 입력받거나 여러 개의 구문구조 중에서 하나를 선택하도록 함으로써 사용자로 하여금 구문구조의 표준에 일치하도록 유도한다.

<149> 그리고 나서, 문장의 구문 분석 결과를 입력받아 의미구조를 생성하여 의미구조 표준에 불일치하거나 의미구조의 중의성이 발생한 경우에 사용자에게 여러 개의 의미구조를 제시하고 사용자로 하여금 하나를 선택하도록 함으로써 사용자의 입력 문장이 의미구조의 표준에 일치하도록 유도하며, 문장의 의미 분석 결과 생성된 의미구조를 의미 관계 노드를 중심으로 분할 및 저장하여 사용자의 정보가 정보 제공을 위한 것일 경우, 의미 관계 노드와 연결되어 있는 개념 노드들의 쌍과 문서의 식별자로 구성된 레코드로 색인 및 저장한다.

<150> 사용자의 정보가 정보 요구를 위한 명세일 경우, 사용자 질의문의 의미구조와 저장되어 있는 제공 정보에 대한 의미구조와의 의미 적합도를 계산하여 의미적으로 가장 근접한 제공 정보를 검색하고, 분석 실패 처리부와 의미구조 처리부로부터 처리 결과를 입력으로 받아 사용자에게 제시할 결과의 형식과 내용을 결정하여 사용자에게 제시할 응답을 생성하여 제시 응답을 출력한다.

<151> 즉, 본 발명의 정보 거래를 위한 정보 생성 및 검색 장치가 정보 제공 측면에서 사용자의 입력 문장을 상기 (1)구문구조 표준과 (2)의미구조 표준에 따라 일치시켜 나가면서 최종적으로 개념 그래프로 색인하는 과정과 정보 요구 측면에서 사용자의 질의어를 분석하여 분할된 개념 그래프에 의해 의미적합도를 계산하여 의미적으로 가장 근접한 정보를 추출해내는 과정을 '좋은 남자용 향수를 파리에서 살 수 있는 방법'이라는 문장을 일실시예로 삼아 설명하도록 한다.

<152> (A) 구문구조 오류 유형 데이터

<153> (0, 306,310)

<154> 오류 유형 0: 표준 위배 오류

<155> (오류타입번호,(표준위배식별자)+)

<156> 오류 유형 1: 구문구조 중의성 발생

<157> (오류타입번호,(구문구조)+)

<158> (B) 사용자 제시 구문구조 유형

<159> (((좋은 (남자용 (향수))를) (파리에서) (((살) 수) 있는) 방법)

<160> (((좋은 남자용) 향수)를) (파리에서) (((살) 수) 있는) 방법)(X)

<161> 상기와 같은 문장 '좋은 남자용 향수를 파리에서 살 수 있는 방법'은 상기 형태소 분석 과정을 수행했을 경우, 형태소 분석 오류가 발생하지 않고 성공적으로 형태소 분석 결과가 생성된다.

<162> 그러나, 상기 구문 분석 과정을 거치면서 상기 구문구조 표준의 속격의 논항 확대와 관형절의 논항 확대 제약 규칙 및 상기 용언 격 프레임 정보에 의해 상기 문장은 분석 실패 처리기(141)에 의해 (207)과정을 통해 상기 (A)와 같은 구문구조 오류 유형 데이터가 생성된다.

<163> 상기 (A)는 상기 문장에 대한 오류 유형 데이터를 예시하고 오류 유형 데이터의 일반적인 데이터 형식(format)을 나타내고 있다. 상기 오류 유형 데이터를 응답 생성기(142)가 입력받아 정보 거래 시스템의 응답 생성 과정(215)을 거쳐 상기 (B)와 같은 결과를 사용자에게 제시하여 사용자로 하여금 문장 구조 중의 하나를 선택하게 하거나 다

시 처음부터 문장을 작성하도록 한다.

<164> 사용자가 선택하거나 구문구조 표준에 일치하여 입력된 문장의 구문구조 트리는 의미구조 생성기(133)에 의해 의미구조로 변환된다.

<165> 도 3 은 도 2a 및 도 2b의 정보 생성/검색 방법 중 의미구조 생성 과정에 대한 일 실시예 상세흐름도이다.

<166> 도 3에 도시된 바와 같이, 의미구조 생성 과정은, 먼저 의미구조 생성기(133)로부터 구문구조 중의성이 해소된 구문 트리(T)를 입력받아(501) 이를 의미적 중의성이 해소되지 않은 개념 그래프인 전단계 개념 그래프(P-CG : Pre-Conceptual Graph)로 변환하는데(502), 이것은 구문 트리의 노드를 재귀적으로 좌우부모노드(left-right-head node) 순으로 탐색해나가면서 좌우 자식 노드로부터 생성된 서브 그래프(subgraph)를 결합하는 방식으로 진행된다. 좌우 자식 노드가 없는 0-레벨 넌-터미널(non-terminal)의 경우, 트리 변환 규칙(307)을 이용하여 서브그래프(subgraph)로 변환한다. 상기 트리 변환 규칙(307)은 도 4a에 도시되어 있는데, 상기 구문 분석기(parser)에서 사용하는 각각의 구문 범주와 그에 따른 변환 규칙을 보여주고 있다.

<167> 이와 같이, 생성된 전단계 개념 그래프(P-CG)(303)는 의미 중의성이 해소된 개념 그래프로 변환되기 위해 개념 노드 처리 과정(304)과 개념 노드간 관계 결정과정(305)을 거치게 된다.

<168> 여기서, 개념 노드의 처리 과정(304)은 수량사, 한정사 처리 규칙(308)을 사용하여 지시어(referent)로 처리되어야 할 정보를 전단계 개념 그래프(P-CG)로부터 찾아서 지시어로 지정하고 개념 분류망(508) 정보를 이용하여 고유명사 및 시제와 같은 내용을 그

개념의 타입 정보로 설정하게 된다. 상기 예문 '좋은 남자용 향수를 파리에서 살 수 있는 방법'에서 '파리' 개념 노드에 대해 개념 처리 결과 생성된 서브그래프가 도 4b에 도시되어 있다. 이어서, 개념 노드 처리가 수행되고 나면 그 개념 노드들 간의 관계가 개념 분류망(309)과 격틀 정보(310)에 의해 결정된다(305). 술어 개념(predicate concept)(단, 동사 및 형용사에 해당함)과 그 술어 개념의 하위 범주(subcategorization)로 들어가는 개념 노드와의 관계는 용언 격틀 정보와 격조사를 이용하여 결정되며, 복합 명사구도 명사구 내의 명사들의 관계를 규칙과 통계적 방법을 사용하여 분석된다.

<169> 이러한 수행 과정 중에는 의미적 중의성이 발생하게 되는데 통계적인 단어 의미 중의성 해결(Word Sense Disambiguation)을 하거나 사용자에게 의미를 선택하도록 여러가지 대안을 제시한다. 예를 들어, 상기 예문 '좋은 남자용 향수를 파리에서 살 수 있는 방법'에서 '파리'와 '사'가 각각 [도시:파리], [곤충:파리] 또는 [사(live)], [사(buy)]의 의미 중의성을 갖게 되는데 사용자에게 두 가지 의미 중에서 어떤 것을 사용한 것인지 묻게 된다. 상기 과정을 수행한 후에, 최종 의미구조인 개념 그래프(306)가 생성된다.

<170> (C) 구문구조 트리 : 좋은 남자용 향수를 파리에서 살 수 있는 방법

<171> (((좋은 ((남자용) 향수))를) ((파리)에서) (((((사)ㄴ) 수) 있)는) 방법)

<172> (D) 의미구조 개념 그래프 : 좋은 남자용 향수를 파리에서 살 수 있는 방법

<173> [방법]→(ATTR)→[Assertion030]

<174> Assertion030: [사]→(AGENT)→[NULL]

- <175> → (OBJECT)→[향수:X]→[ATTR][Assertion:031]
- <176> →(JUXTAPOSE)[남자용]
- <177> → (LOCATE)→[파리]
- <178> → (ATTR)→[값싸]
- <179> Assertion031 : [좋]→(EXPERIENCER)→[X]
- <180> (E) 의미구조 오류 유형 데이터
- <181> (1, 파리, 파리/도시, 파리/곤충, 사, 사/live, 사/buy)
- <182> 의미구조 오류 유형 0: 표준 의미구조 위배
- <183> (0, (표준의미구조 위배 항목)+)
- <184> 의미구조 오류 유형 1: 의미 중의성 발생
- <185> (1, 의미 중의성 발생 노드 식별자, (의미 중의성 발생 노드 식별자의 의미)++)
- <186> (F) 분할 그래프
- <187> (ATTR)[방법][Assertion030]
- <188> (AGENT)[사][NULL]
- <189> (OBJECT)[사][향수:X]
- <190> (LOCATE)[사][파리]
- <191> (ATTR)[사][값싸]
- <192> (ATTR)[X][Assertion031]
- <193> (JUXTAPOSE)[X][남자용]

<194> (EXPERIENCER)[종][X]

<195> 상기 (C) 및 (D)는 상기 예문 '좋은 남자 향수를 파리에서 살 수 있는 방법'에 대한 상기 의미구조 생성 과정(208)에 대한 입력으로서, 구문 트리와 출력으로 의미구조인 개념 그래프를 예시한다. 상기 (D)에서 예시된 의미구조에서는 동사에 대한 시제나 상 정보와 같은 것을 편의상 생략했다.

<196> 상기 과정을 거쳐 최종적으로 생성된 의미구조에 대해서는 상기 (2)의 의미구조 표준에 적합한지의 여부를 판별한다. 적합 여부의 판별은 사용자에게 최종적으로 생성된 개념 그래프를 제시하여 사용자가 의도하는 의미와 일치하는지를 확인함으로써 이루어진다. 상기 (E)는 의미구조 표준에 적합하지 않거나 의미 중의성이 발생할 경우의 의미구조 생성 오류 유형 데이터(210)에 대한 예와 의미구조 오류 유형 데이터 형식을 보여준다. 사용자에게 의해 확인된 의미구조가 의미구조 표준에 적합하면 그 개념 그래프를 관계에 따라 분할한다(211). 상기 (D)에 예시된 개념 그래프를 관계 노드에 따라 분할한 결과 생성된 분할 그래프가 상기 (F)에 예시되어 있다.

<197> 개념 그래프가 분할된 이후엔 현재의 사용자의 입력 문장이 정보 제공을 위한 것인지 아니면 정보 요구를 위한 질의어인지에 따라, 개념 그래프의 색인이나 개념 그래프의 검색이 이루어지게 된다. 정보 제공일 경우, 수행되는 개념 그래프 색인은 관계에 대한 개념 쌍들과 웹 문서 식별자로 이루어지는 레코드들로 구성되는 관계형 데이터베이스의 테이블을 생성하는 과정이다. 상기 (F)에 예시된 분할 그래프들을 색인한 결과 생성되는 테이블 구조와 내용이 도 5a 내지 도 5f에 예시되어 있다. 분할된 그래프가 사용자의 요구 정보를 나타내는 질의어에 대한 것일 경우에는 개념 그래프 검색을 수행하여 질의어

를 통해 나타나는 사용자의 요구 정보와 의미적으로 가장 근접한 문서를 추출하게 된다.

<198> 도 6 은 본 발명에 따른 사용자의 요구 정보 및 문서 추출 과정에 대한 일실시에 흐름도이다.

<199> 도 6에 도시된 바와 같이, 먼저 초기화 과정($d \leftarrow$ 최상위 노드의 레벨=0, $N \leftarrow$ 분할 그래프의 깊이(depth))을 거친 후(901)에 최상위 노드의 레벨과 분할 그래프의 깊이를 비교하여(602) 분할 그래프의 깊이 보다 작으면 개념 그래프의 레벨(d)에 속하는 관계 노드에 대해 탐색을 시작한다(603).

<200> 이어서, 언어 특성적 탐색 우선순위 노드($c1$, $c2$)를 결정하고(604), $c1$ 과 $c2$ 에 대해서 언어 $L1$ 의 우선 규칙(608)에 따라, 관계 노드(n)과 연관된 테이블의 레코드를 탐색하여(605) 각 레코드의 의미 적합도를 계산하면서(606) 최상의 노드의 레벨을 증가시킨 후(607), 최상위 노드의 레벨과 분할 그래프의 깊이를 비교하는 과정(602)을 반복 수행한다.

<201> 즉, 초기화 과정($d \leftarrow$ 최상위 노드의 레벨=0, $N \leftarrow$ 분할 그래프의 깊이(depth))을 거친 후에, 아직 탐색해야 할 노드가 남아 있을 시에는(602), 현재의 그래프 레벨에 속하는 관계노드에 대한 탐색을 시작하게 된다. 그래프 레벨은 개념 그래프의 계층 구조의 위치를 의미한다.

<202> 상기 (D)에 예시된 개념 그래프와 상기 (D)에 예시된 분할 그래프에서와 같이 도 5a의 'ATTR' 테이블은 최상위 레벨의 관계 노드이며, 도 5b의 'AGENT' 테이블은 차상위 레벨의 관계노드가 된다. 개념 그래프의 깊이(depth)는 그래프의 단말 노드까지의 레벨의 총 수가 되며, 상기 (D)에 예시된 개념 그래프의 깊이(depth)는 4가 된다.

<203> 탐색을 수행할 관계 노드가 결정되면 그 관계 노드에 연결된 한 쌍의 개념 노드들에 대해서 언어 특성적인 의미 적합 우선순위를 결정한다. 의미 적합 우선순위는 의미가 첨가되는 수식+피수식 관계에 있을 때 피수식어에 해당하는 개념 노드가 우선순위를 갖는다. 그 이외의 경우에는 언어에 고유하게 존재하는 의미 우선 순위 규칙에 따라 결정을 하게 된다. 우선순위가 결정되고 나면 각 개념 노드에 대하여 관계 노드 n에 대한 테이블의 각 레코드와의 의미 적합도를 계산한다. 개념 x와 y 사이의 의미 적합도 $S(x,y)$ 는 다음의 (수학식 1)에 의해 계산된다.

<204> 【수학식 1】

$$S(x,y) = \frac{1}{1+d(x,y)}, \text{ where } d(x,y) \text{ 개념망에서의 노드간 거리}$$

<205> $d(x,y)$, 즉 개념망에서 노드 x에 대한 y의 거리는 y가 x의 하위 노드중의 하나일 경우에는 0이 되며, 그 이외의 경우에는 노드간 에지의 수로 계산된다. 예를 들어 $d('방법', '절차')$ 의 값은 만일 개념망 내에서 '절차'가 '방법'의 하위 개념일 때 0의 값을 갖게 된다. 따라서, 웹문서(I)가 특정 질의어에 대해 갖는 의미 적합도 $S(I,Q)$ 는 I가 갖는 모든 문장의 개념 그래프에 대해 모든 테이블내에서 I에 대해 색인되어 있는 레코드에 대한 의미적합도의 합으로 계산되며 그것은 (수학식 2)와 같이 표현된다.

<206> 【수학식 2】

$$S(I,Q) = \sum_l^L \sum_m^M \sum_n^N \frac{1}{d} \{ \delta \cdot s(c_1^l, r_1^{mn}) + (1-\delta) \cdot s(c_2^l, r_2^{mn}) \}$$

- <207> 개념 그래프 검색이 수행된 후에는 검색된 그래프에 대해 S(I,Q) 값의 순서대로 정렬을 하고, 그 결과를 응답 생성기(142)에 전달한다. 응답 생성기(142)는 색인 결과나 검색 결과 혹은 분석 오류 유형 데이터를 입력받아서 사용자에게 응답 문장 및 결과 출력 양식을 생성하고 정보 제공 인터페이스(16)나 정보 수요 인터페이스(12)를 통해 사용자에게 출력된다.
- <208> 상술한 바와 같은 본 발명의 방법은 프로그램으로 구현되어 컴퓨터로 읽을 수 있는 기록매체(씨디롬, 램, 롬, 플로피 디스크, 하드 디스크, 광자기 디스크 등)에 저장될 수 있다.
- <209> 이상에서 설명한 본 발명은 전술한 실시예 및 첨부된 도면에 의해 한정되는 것이 아니고, 본 발명의 기술적 사상을 벗어나지 않는 범위 내에서 여러 가지 치환, 변형 및 변경이 가능하다는 것이 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 있어 명백할 것이다.

【발명의 효과】

- <210> 상기한 바와 같은 본 발명은, 영어나 한국어와 같은 자연언어 문장을 통해 제공하고자 하는 정보나 얻고자 하는 정보를 정보거래를 위한 구문구조 표준과 의미구조 표준에 맞추어 기술(description)함으로써, 인터넷에서 폭발적으로 늘어나는 정보의 효율적인 검색을 가능하게 하고, 구문구조와 의미구조의 표준에 따르도록 정보 생성 및 작성 단계에서부터 유도함으로써 인터넷에서 정확하고 명확한 정보의 생성을 가능하게 하여

정보의 거래 및 유통을 활성화시킬 수 있는 효과가 있다.

<211> 또한, 본 발명은, 사용자의 정보 요구를 표현하는 질의어나 제공 정보를 나타내는 문장에 대해 표층적인 구문구조 분석과 심층적인 의미구조 분석 및 의미구조인 개념그래프의 의미 적합도 계산을 통해 의미적으로는 동일하지만 표층 구조가 다른 문장으로 이루어진 정보에 대해서도 검색이 가능하게 함으로써, 사용자가 필요로 하는 정보에 대한 요구나 사용자가 제공하고자 하는 정보의 거래에 대한 요구를 충족시킬 수 있는 효과가 있다.

【특허청구범위】**【청구항 1】**

표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치에 있어서,
사용자로부터의 정보 제공을 위한 문장이나 정보 요구를 위한 질의문을 분석하기
위해 사용하는 언어 지식 데이터와 문장의 의미를 개념 그래프로 나타내는 의미 표현 데
이터와 웹 문서를 저장하기 위한 데이터 저장수단;

상기 사용자로부터 정보 제공을 위한 자연언어 문장의 생성이나 정보의 요구 명세
를 위한 자연어 질의 문장을 입력받기 위한 입력수단;

상기 데이터 저장수단의 데이터를 참조하여 상기 사용자로부터 입력된 자연언어
문장 또는 자연언어 질의 문장의 구문구조를 분석하여 의미구조를 생성하기 위한 입력
문장 분석수단;

상기 입력 문장 분석수단을 통해 분석된 의미 구조를 분할하여 데이터베이스에 색
인하여 저장하거나 의미 적합도를 계산하여 요구 정보 명세에 가장 의미적으로 적합한
제공 정보 및 문서를 검색하기 위한 의미구조 처리수단;

상기 입력 문장 분석수단으로부터 제시된 오류 데이터를 구문구조 표준과 의미구
조 표준에 따라 수정한 문장 형식 규칙, 상기 색인/검색 결과를 출력하기 위한 대화형
처리수단; 및

상기 대화형 처리수단으로부터 출력되는 데이터를 사용자에게 제공하기 위한 정보
제공수단

을 포함하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치.

【청구항 2】

제 1 항에 있어서,

상기 입력 문장 분석수단은,

상기 사용자로부터 입력된 문장을 전달받아 데이터 저장부에 있는 사전 저장 장치, 용언 격 프레임 저장 장치 및 명사 개념 분류망 저장 장치의 데이터와 비교하여 순차적으로 분석하고, 형태소 분석기에서 형태소를 분석하며, 구문 분석기에서 구문을 분석하여 구문구조 트리를 생성하고, 의미구조 생성기에서 의미 분석을 수행하여 의미구조를 생성하는 것을 특징으로 하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치.

【청구항 3】

제 1 항 또는 제 2 항에 있어서,

상기 의미구조 처리수단은,

상기 의미구조 생성기로부터 출력된 개념 그래프를 의미 관계에 따라 변환하기 위한 개념 그래프 변환기;

상기 개념 그래프 변환기를 통해 변환된 각각의 관계로 연결된 개념 쌍의 레코드에 의해 사용자의 제공 정보가 있는 원래의 웹문서에 대해 색인을 수행하기 위한 개념 그래프 색인기; 및

사용자 질의어의 의미구조와 저장된 의미구조들 사이에서 의미 적합도(semantic relevance)를 계산하여 적합도가 높은 제공 정보를 탐색하기 위한 개념 그래프 검색기를 포함하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치.

【청구항 4】

제 3 항에 있어서,

상기 대화형 처리수단은,

사용자가 작성한 문장에서 발생하는 철자나 띄어쓰기 오류, 구문구조나 의미구조의 표준과 위배되는 비문, 구문구조의 중의성 및 의미구조의 중의성 등의 분석 실패를 해결하는 것을 특징으로 하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치.

【청구항 5】

표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치에 적용되는 정보 생성/검색 방법에 있어서,

정보 공급자에 의해 기술되는 자연언어 문장(정보 및 지식)을 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환하여 색인하는 제 1 단계; 및

사용자로부터 입력되는 자연언어 질의어 문장을 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환하여 상기 색인된 정보중에서 사용자의 요구사항에 적합한 정보를 검색하여 제공하는 제 2 단계

를 포함하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법.

【청구항 6】

제 5 항에 있어서,

상기 제 1 단계는,

상기 정보 공급자에 의해 입력되는 자연언어 문장을 구문구조 및 의미구조의 표준에 따라 구문구조 및 의미구조의 중의성이 해소된 문장으로 생성하는 제 3 단계;

상기 생성된 문장을 구문분석과 의미분석을 통하여 개념 그래프로 변환하는 제 4 단계; 및

상기 변환된 개념 그래프를 관계 노드에 의한 테이블의 레코드 값으로 변환하여 색인하는 제 5 단계

를 포함하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법.

【청구항 7】

제 5 항 또는 제 6 항에 있어서,

상기 제 1 단계는,

정보 거래를 위한 정보를 생성하여 표준에 적합한 구조로 사용자의 문장을 유도하여 구문구조 표준과 의미구조 표준에 적합한 문장을 생성하는 것을 특징으로 하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법.

【청구항 8】

제 7 항에 있어서,

상기 제 1 단계는,

상기 사용자의 문장을 유도하여 구문구조 분석과 의미구조 생성 실패시, 사용자로부터 올바른 구문이나 의미를 선택하도록 여러가지 분석 결과를 분석 순위에 따라 사용자에게 제시하는 것을 특징으로 하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법.

【청구항 9】

제 5 항에 있어서,

상기 제 2 단계는,

상기 사용자로부터 자연언어 질의어 문장을 입력받아 구문구조 및 의미구조를 분석하여 개념 그래프로 변환하는 제 3 단계;

상기 질의어를 개념 그래프와 의미적으로 가장 근접한 데이터베이스내의 개념 그래프를 탐색하여 의미 적합도를 계산하는 제 4 단계; 및

상기 탐색된 개념 그래프에 의해 색인된 정보를 추출하여 상기 사용자에게 제공하는 제 5 단계

를 포함하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법.

【청구항 10】

제 5 항, 제 6 항 또는 제 9 항 중 어느 한 항에 있어서,

상기 정보 공급자에 의해 기술되는 자연언어 문장(정보 및 지식) 및 상기 사용자로부터 입력되는 자연언어 질의어 문장이 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환되는 과정은,

상기 정보 공급자 자신이 제공하고자 하는 정보나 제공받기를 원하는 정보에 대한 자연어 문장이 입력되면 형태소 분석기를 통하여 자연어 문장에 대한 형태소를 분석하여 성공 여부를 확인하는 제 6 단계;

상기 제 6 단계의 확인결과, 형태소 분석에 실패하였으면 오류가 발생한 유형에 따라 오류 유형 데이터를 생성하고, 형태소 분석에 성공하였으면 형태소 분석 결과를 가지고 구문구조를 분석하는 제 7 단계;

상기 분석된 구문구조의 생성 여부에 따라 구문 분석 트리를 의미구조로 변환하여 의미구조를 생성하는 제 8 단계; 및

상기 의미구조가 표준에 적합하게 생성되었는지의 여부에 따라 생성된 의미구조를 개념 그래프 변환기로 입력하여 상기 개념 그래프에 대한 분할을 수행하는 제 9 단계

를 포함하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법.

【청구항 11】

제 10 항에 있어서,

상기 의미구조 생성 과정은,

의미구조 생성기로부터 구문구조 중의성이 해소된 구문 트리(T)를 입력받아 구문 트리(T)를 트리 변환 규칙에 따라 전단계 개념 그래프(P-CG)로 변환하는 제 10 단계;

상기 변환된 전단계 개념 그래프(P-CG)를 의미 중의성이 해소된 개념 그래프로 변환하기 위해, 수량사 및 한정사 처리 규칙을 이용하여 지시어로 처리되어야 할 정보를 상기 전단계 개념 그래프(P-CG)로부터 찾아서 지시어로 지정하고 개념 분류망 정보를 이용하여 고유명사 및 시제와 같은 내용을 그 개념의 타입 정보로 설정하여 개념 노드를 처리하는 제 11 단계; 및

상기 개념 노드 처리 수행 후, 상기 개념 분류망 정보와 격률 정보에 의해 개념 노드들 간의 관계를 결정하여 최종 의미구조인 개념 그래프를 생성하는 제 12 단계

를 포함하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법.

【청구항 12】

제 10 항에 있어서,

상기 개념 그래프 분할 수행 과정은,

상기 정보 공급자의 요구 정보 및 문서를 추출하기 위하여, 최상위 노드 레벨(d), 분할 그래프의 깊이(N)를 초기화시키는 제 10 단계;

상기 초기화 수행 후, 최상위 노드의 레벨(d)과 분할 그래프의 깊이(N)를 비교한 결과에 따라, 상기 개념 그래프의 레벨(d)에 속하는 관계 노드(n)에 대해 탐색하는 제 11 단계;

언어 특성적 탐색 우선순위 노드(c1, c2)를 결정하고, 결정된 우선순위 노드(c1,

c2)에 대해서 언어(L1~Ln)의 우선순위 규칙에 따라, 관계 노드(n)와 연관된 테이블의 레코드를 탐색하여 각 레코드의 의미 적합도($S(x,y)$)를 계산하는 제 12 단계; 및

상기 의미 적합도($S(x,y)$) 계산 결과에 따라, 최상의 노드의 레벨(d)을 증가시킨 후, 상기 제 10 단계로 진행하는 제 13 단계

를 포함하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법.

【청구항 13】

제 12 항에 있어서,

상기 의미 적합도($S(x,y)$)는,

하기의 수식과 같이 개념망에서 노드 x에 대한 y의 거리($d(x,y)$)를 나타내는 것을 특징으로 하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법.

$$S(x,y) = \frac{1}{1+d(x,y)}, \text{ where } d(x,y) \text{ 개념망에서의 노드간 거리}$$

(여기서, 개념망에서 노드 x에 대한 y의 거리($d(x,y)$)는 y가 x의 하위 노드중의 하나일 경우에는 0이 되고, 그 이외의 경우에는 노드간 에지의 수로 계산됨)

【청구항 14】

제 5 항 또는 제 9 항에 있어서,

상기 제 2 단계는,

의미구조 그래프를 분할하여 의미 관계를 중심으로 정보를 저장하여 검색하고, 명

사 개념 분류망을 이용해 개념과 개념간의 의미 적합도를 이용하여 사용자의 요구 정보 명세에 가장 근사한 정보와 문서를 검색하는 것을 특징으로 하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 방법.

【청구항 15】

표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성 장치에 적용되는 정보 생성 방법에 있어서,

상기 정보 공급자에 의해 입력되는 자연언어 문장을 구문구조 및 의미구조의 표준에 따라 구문구조 및 의미구조의 중의성이 해소된 문장으로 생성하는 제 1 단계;

상기 생성된 문장을 구문분석과 의미분석을 통하여 개념 그래프로 변환하는 제 2 단계; 및

상기 변환된 개념 그래프를 관계 노드에 의한 테이블의 레코드 값으로 변환하여 색인하는 제 3 단계

를 포함하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성 방법.

【청구항 16】

제 15 항에 있어서,

상기 정보 생성 방법은,

정보 거래를 위한 정보를 생성하여 표준에 적합한 구조로 사용자 문장을 유도하여

구문구조 표준과 의미구조 표준에 적합한 문장을 생성하는 것을 특징으로 하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성 방법.

【청구항 17】

제 15 항 또는 제 16 항에 있어서,

상기 정보 생성 방법은,

상기 사용자의 문장을 유도하여 구문구조 분석과 의미구조 생성 실패시, 사용자에게 제시하는 것을 특징으로 하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성 방법.

【청구항 18】

표준화된 문장 구문구조 및 의미구조에 기반한 정보 검색 장치에 적용되는 정보 검색 방법에 있어서,

상기 사용자로부터 자연언어 질의어 문장을 입력받아 구문구조 및 의미구조를 분석하여 개념 그래프로 변환하는 제 1 단계;

상기 질의어를 개념 그래프와 의미적으로 가장 근접한 데이터베이스내의 개념 그래프를 탐색하여 의미 적합도를 계산하는 제 2 단계; 및

상기 탐색된 개념 그래프에 의해 색인된 정보를 추출하여 상기 사용자에게 제공하는 제 3 단계

를 포함하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 검색 방법.

【청구항 19】

제 18 항에 있어서,

상기 정보 검색 방법은,

의미구조 그래프를 분할하여 의미 관계를 중심으로 정보를 검색하고, 명사 개념 분류망을 이용해 개념과 개념간의 의미 적합도를 계산하여 사용자의 요구 정보 명세에 가장 근사한 정보와 문서를 검색하는 것을 특징으로 하는 표준화된 문장 구문구조 및 의미구조에 기반한 정보 검색 방법.

【청구항 20】

프로세서를 구비한 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성/검색 장치에,

정보 공급자에 의해 기술되는 자연언어 문장(정보 및 지식)을 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환하여 색인하는 제 1 기능; 및

사용자로부터 입력되는 자연언어 질의어 문장을 구문구조 및 의미구조 표준에 따라 개념 그래프로 변환하여 상기 색인된 정보중에서 사용자의 요구사항에 적합한 정보를 검색하여 제공하는 제 2 기능

을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

【청구항 21】

프로세서를 구비한 표준화된 문장 구문구조 및 의미구조에 기반한 정보 생성 장치에,

상기 정보 공급자에 의해 입력되는 자연언어 문장을 구문구조 및 의미구조의 표준에 따라 구문구조 및 의미구조의 중의성이 해소된 문장으로 생성하는 제 1 기능;

상기 생성된 문장을 구문분석과 의미분석을 통하여 개념 그래프로 변환하는 제 2 기능; 및

상기 변환된 개념 그래프를 관계 노드에 의한 테이블의 레코드 값으로 변환하여 색인하는 제 3 기능

을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

【청구항 22】

프로세서를 구비한 표준화된 문장 구문구조 및 의미구조에 기반한 정보 검색 장치에,

상기 사용자로부터 자연언어 질의어 문장을 입력받아 구문구조 및 의미구조를 분석하여 개념 그래프로 변환하는 제 1 기능;

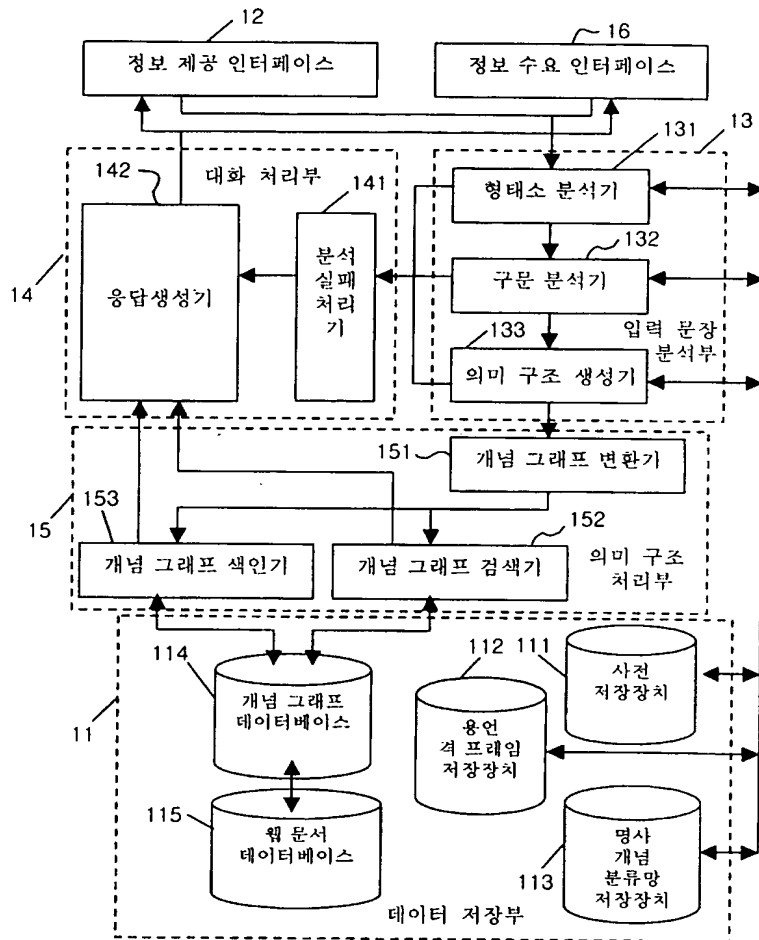
상기 질의어를 개념 그래프와 의미적으로 가장 근접한 데이터베이스내의 개념 그래프를 탐색하여 의미 적합도를 계산하는 제 2 기능; 및

상기 탐색된 개념 그래프에 의해 색인된 정보를 추출하여 상기 사용자에게 제공하는 제 3 기능

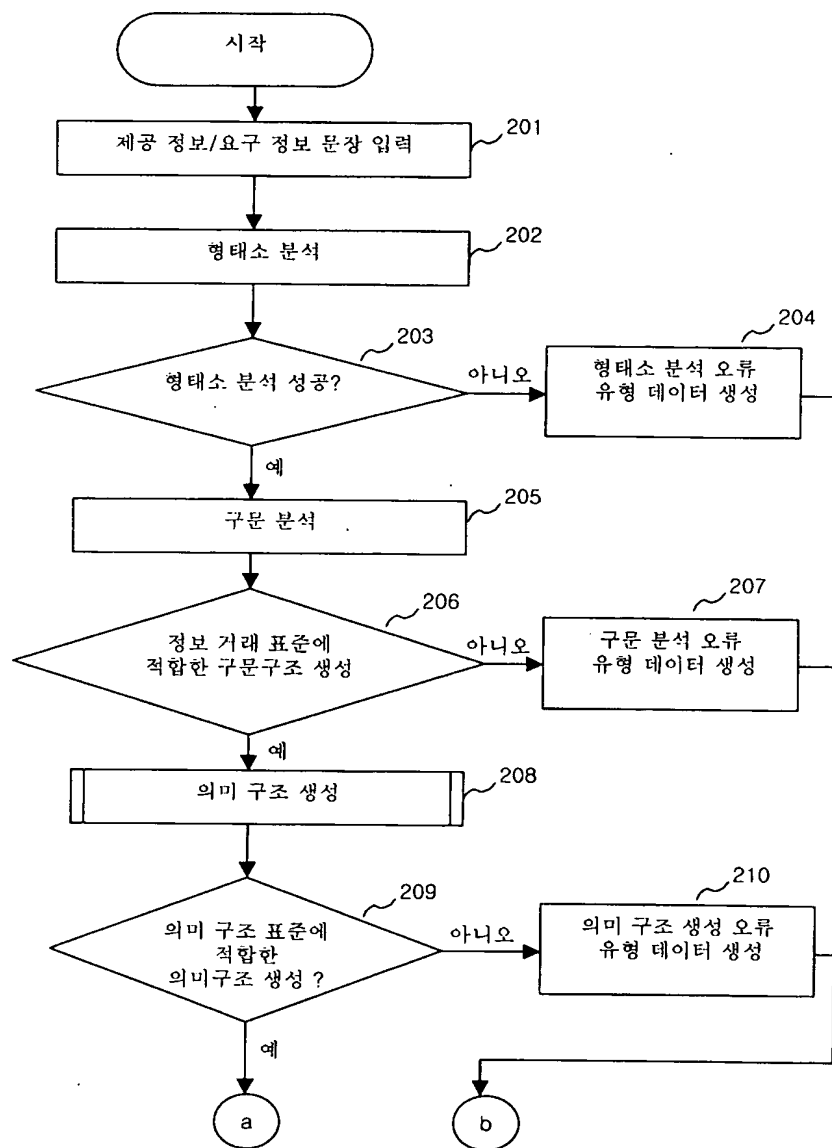
을 실현시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

【도면】

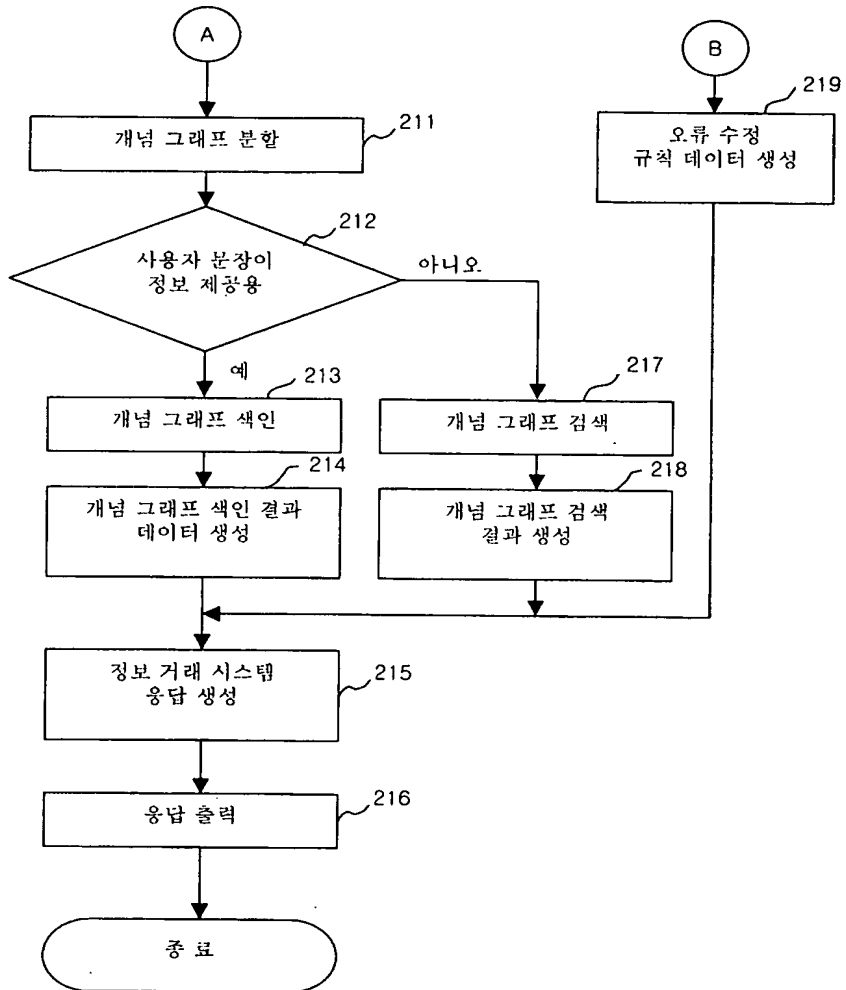
【도 1】



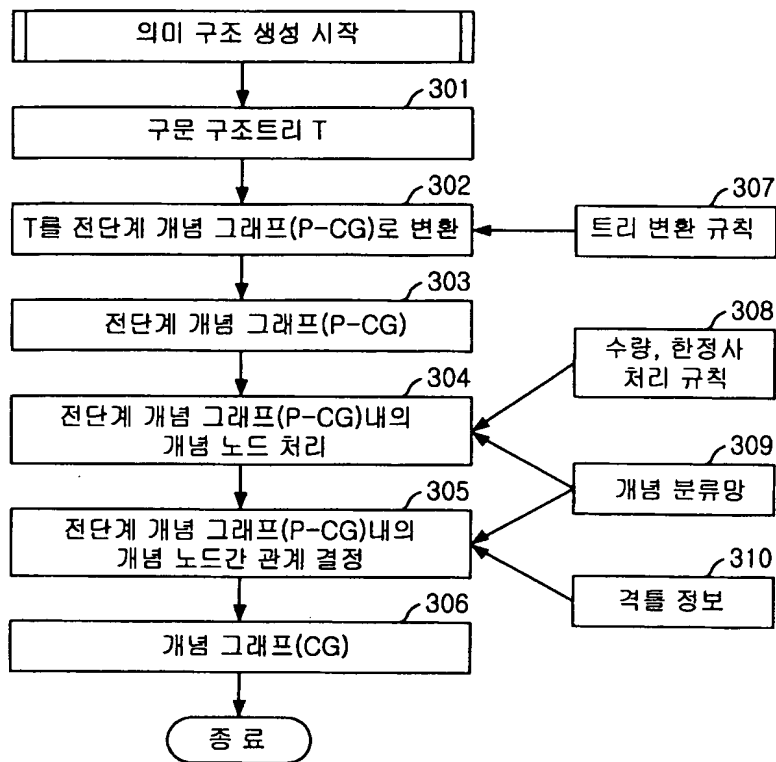
【도 2a】



【도 2b】



【도 3】



【도 4a】

구문 범주	변환 규칙
V-, C-, D-, n-	관계 노드만을 포함한 서브그래프 생성
N0, V0	개념 노드만을 포함한 서브그래프 생성
D0, A0	개념 노드 및 '수식' 관계 노드를 연결한 서브그래프 생성

【도 4b】

개념 노드 처리 이전	개념 노드 처리 이후
[파리/NUP]	[국가:프랑스]→(PART_OF)→[도시: 파리]

【도 5a】

ATTR 테이블

개념 1	개념 2	문서 아이디
방법	Assertion030	1
항수X	Assertion031	1
사/buy	값싸	1

【도 5b】

AGENT 테이블

개념 1	개념 2	문서 아이디
사/buy	NULL	1
:	:	:

【도 5c】

EXPERIENCER 테이블

개념 1	개념 2	문서 아이디
중/good	항수X	1
:	:	:

【도 5d】

OBJECT 테이블

개념 1	개념 2	문서 아이디
사/buy	항수X	1
:	:	:

【도 5e】

LOCATION 테이블

개념 1	개념 2	문서 아이디
사/buy	파리	1
:	:	:

【도 5f】

JUXTAPOSE 테이블

개념 1	개념 2	문서 아이디
항수X	남자용	1
:	:	:

【도 6】

